

DOCUMENT RESUME

ED 061 023

RE 004 074

AUTHOR Jackson, Rex
TITLE The Development of a Collection of Adult Tasks for Assessment of Literacy.
PUB DATE Apr 72
NOTE 10p.; Paper presented at the meeting of the American Educational Research Association, Chicago, Ill., Apr. 1972

EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS *Adult Literacy; *Criterion Referenced Tests; Item Analysis; Literacy; Literacy Education; *Reading Skills; Research and Development Centers; *Task Analysis; *Test Construction

ABSTRACT

Efforts to develop a collection of reading test items suitable for measuring adult literacy are described. The project, part of Project I of the Targeted Research and Development Program in Reading, sought to identify characteristics of literacy levels and to define tasks which might be used in a criterion-referenced test to assess possession of those characteristics. It was felt that such tasks should realistically reflect adults' everyday reading behavior and should be associated with benefits which adults might realize from being able to perform the tasks. It was also decided that tasks should be practical rather than academic and that numbers of tasks of a given point be based on frequency and duration of time spent in real-life activities upon which the tasks are based. In addition, self-reported nonreaders' statements of their abilities ought to be considered. It is planned that once items are developed based on such criteria, a panel of test developers will construct psychometrically sound instruments from the large bank of items developed by project participants. (MS)

ED 061023

The Development of A Collection
of Adult Tasks for
Assessment of Literacy

Rex Jackson
Yale University

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

For Presentation at the Annual Meeting
of the American Educational
Research Association
Chicago, Illinois
April 4, 1972

RE 004 074

The Development of a Collection of
Adult Tasks for Assessment of Literacy

Rex Jackson¹
Yale University

This paper is principally a report on some of the thinking underlying task development procedures used in Project I of the Targeted Research and Development Program in Reading - a project currently being conducted by Educational Testing Service under sponsorship of the U.S. Office of Education. The principal objective of this project is the development of a collection of reading tests which will be available as a resource in the construction of criterion instruments for assessing the effectiveness of large-scale, national efforts to raise literacy levels among adults. It should be noted that the product will be a collection - an item bank - rather than a "test" with closely specified psychometric properties, since this feature has had a major bearing on the design of the project. The basic objective is not development of an instrument used to assign individuals to one or two categories - literate and non-literate - in a meaningful way, but rather the creation of a pool of tasks from which different subsets of tasks could be culled to serve a variety of assessment purposes. At this stage the field trials of the tasks have not been performed. Therefore, this report is limited to the initial task development and selection phases of the work.

Perhaps the principal feature that distinguishes this project from other efforts to develop measures of functional literacy is its scope.

¹

The author's participation in the project described in this paper occurred while he was a member of the staff of Educational Testing Service.

A number of such efforts have been concerned with assessing reading ability as it relates to specific functional competencies of rather narrowly defined target populations. An example would be the development of special purpose instruments to be used in assessing whether individuals have the reading skills prerequisite to success in particular occupational specialties. Partly because of their narrow focus, such instruments can be quite valuable in assisting specific selection and placement decisions. In contrast to such efforts, however, the Targeted Rand D Project is concerned with reading as it affects the abilities of members of a highly heterogeneous population - virtually all adults in the U.S. - to perform a great variety of both occupational and non occupational roles. The reading tasks developed are concerned not only with reading incidental to occupational or income-producing activities, but also with use of income, consumption, education and self-development, preservation of health, participation in political processes, and recreation - to name only several of the areas of daily life in which reading commonly plays either a central or an important supporting role.

When both the heterogeneity of the adult population and the diversity of reading-related activities in which members of that population engage are considered, it is difficult to rationalize the usual two-state, pass-fail construction of "functional literacy." One must consider individuals and groups with dissimilar characteristics and aspirations existing in environments that present them with differing demands and possibilities of reward. For example, the obvious first partition of the population - into men and women - already yields clusters of social roles (and consequent reading demands) that are much more frequently performed by one

group than the other - at least in our current culture. "Functioning" may be evidenced by different sets of behaviors depending on such personal and environmental variables. Thus, it seems unwise to propose a simple, dichotomous standard applicable to all that can serve as a reference for a normative statement of the type: this individual can read "well enough."

Despite the above, it is evident that the ability to read is integral to many competencies that the society values (and rewards) on the part of its citizens and that reading with some degree of skill is prerequisite to a variety of behaviors that would be included in any reasonable definition of "personal success." Few would dispute the linkage of reading ability and performance of significant life roles suggested by the term "functional literacy." Differences arise, however, with regard to the uniform standard implied by the word "literacy." It appears, therefore, more appropriate to conceptualize "literacy" as a trait capable of continuous measurement. If other relevant variables are held constant, it is reasonable to assume that ability to share in many of the economic and other benefits afforded by the society will increase monotonically as one moves higher on this literacy continuum - at least to some ceiling. It may well be undesirable (as well as unsound) to specify some minimal threshold level to serve either as a target for reading instruction or the basis of a classification rule.

The discussion above is offered with due acknowledgement of the virtues of simplicity and of the considerable convenience afforded by instruments and decision rules that classify individuals into one of two exclusive categories. Such classifications afford both ease in assignment of individuals to alternative treatments and descriptive simplicity.

Percentage statements about the incidence of "functional illiteracy" in the population are likely to have a greater impact on the public and its representatives than more elaborate reports. It is suggested, however, that the ultimately arbitrary definitions that must be made in order to achieve this desired simplicity lead to a great loss of information about the real effects of reading disability on various kinds of individual functioning.

If a more complex model of functional literacy is adopted, then, what principal of selection should guide development of tasks to assess this trait? One approach to answering this question involves consideration of the ways in which empirical validation of tasks might be carried out. It would not be adequate simply to examine the degree to which tasks discriminate among individuals above and below some specified critical level on some uni-dimensional measure of functional competence. Instead interest should focus on the incremental benefits afforded to individuals as levels of success increase. So the test construction question becomes: what tasks are likely to provide effective discrimination with respect to measures of benefit?

It may be argued that items typical of a conventional reading comprehension test would be as suitable as anything else for this purpose. This argument is based on the view that reading is an essentially uni-dimensional trait. Thus any suitably difficult, good measure of this generalized ability - such as standard reading tests of the type commonly used in schools - might well be as valid as tasks specially constructed with benefit criteria in mind. Although this appears to be a reasonable

argument, more specialized reading tasks were developed for this project. Several considerations lend support to the approach actually used: construction of reading items based on non-academic reading materials and paralleling as closely as possible tasks actually encountered in practical settings. Despite evidence suggesting that reading is essentially unitary, it seems possible that such tasks would have marginally greater validity - e.g. that successful performance of a task based on, say, a typical instruction manual would be a better indicator of occupationally related reading skill than ability to answer a multiple choice question about a textbook-like passage. A further desirable feature of practical tasks is their realism or face validity - a factor which should not be undervalued. It seems highly likely that adults will be more highly motivated to perform well on realistic, practical tasks than on academic materials. A third point in favor of practical tasks is the possibility of drawing inferences from performance without reference to such statistics as levels of success among subpopulations. If groups of tasks can be considered "work-samples" from specified domains of activity, one can interpret performance directly with regard to functional competence. (The "if" is important here, however.)

Without slighting the value of empirical validation, then, attempts were made to introduce benefit considerations into the construction of tasks themselves - to identify those "real-life" reading activities that do provide substantial payoffs to those who engage in them. One difficulty in choosing among potential tasks on the basis of this kind of criterion, however, was introduced by the fact that the reading to be represented might have been performed for a variety of purposes. There is no clear-

cut objective way, for example, to compare the value of reading a cookbook recipe with that of reading a road sign. Instead of attempting this kind of dubious comparison, reading tasks performed in the service of different general purposes were considered separately. To a large extent initial selection was a judgemental process with such characteristics considered as the extent to which the information acquired by reading a particular piece of material did in fact facilitate attainment of desired goals, the extent to which the reading was sufficient for this purpose, and the extent to which the reading was necessary, i.e. the only practical alternative.

A number of reading demands are self-evidently both commonly encountered and of critical importance. For example, reading the instructions on a medicine bottle is clearly an important task that will be encountered by most adults a number of times in their lives; the potential adverse consequences of failure to read these instructions correctly are obvious. A great many tasks fall into this category of self-evident high-benefit reading. If one asks (as we did) a number of individuals to list reading tasks that they would classify as being particularly critical or valuable, substantial overlap among the lists is found. There was no shortage of potential tasks competing for a place in the collection. The problem was in choosing among the likely candidates.

The principal systematic information bearing on these selection choices was developed through a national survey designed to yield a comprehensive picture of the reading activities of adults. The basic structure of the survey interviews through which this data was gathered involved having a respondent break down a 24 hour day into general activity segments -

work, travel, meals, shopping, recreation, and the like. He was then asked to report any reading he performed that was associated with those activities during that day. Although this structure was used partly because it was judged that consideration of general activities would prompt recall of specific reading incidents, it also allowed linkage of reading and other activities and thus the drawing of some inferences about the kinds of reading generally engaged in for specific purposes. Separate sections of the interview dealt with purposeful types of reading that are not normally associated with other activities - newspapers, magazines, books, and mail.

In the analysis of the survey data, it was not assumed that there is one single measure of reading benefits derivable from self-reports of reading activities. Instead, for specific classes of reading activities, a number of statistics were calculated which were to be taken as indicators or evidence of benefit. These indicators were regarded as guides to judgement in task selection, not as selection criteria themselves. Among the indicators were:

- 1) Frequencies for reading activities. It is natural to assume that reading that is valuable in a variety of circumstances will be reported with fairly high frequency. Very rare activities would not be good candidates for the collection.
- 2) Duration. Similarly, the time spent in reading activities should have some relationship to benefit, though this is obviously a much weaker indicator than frequency.
- 3) Importance ratings of reading activities by respondents. For each activity, engaged in, respondents were asked directly to report its

importance to them.

- 4) Statements by self-reported non-readers. Those who stated that they could not read and understand English at a specified minimum level were asked to name kinds of reading that they believed would be particularly important or useful to them. These reports were treated as "anecdotal" or "clinical" information rather than as statistical data.

Aside from development of these indicators, analyses were performed relating reading activities and demographic data. Of particular interest here was the identification of reading activities common to subpopulations and of reading activities characteristic of individuals at particular economic, occupational, and educational levels.

As described thus far, then, test selection was principally based on the judgement of professional test developers supplemented by the survey data. In order to broaden the selection basis, however, panels of advisers were assembled to provide additional input with regard to reading benefits. The panels consisted of representatives of such professional areas as consumer and voter education, occupational training, basic adult education, and public health. The intention was to supplement the self-reported "man-on-the-street" information from the survey with the judgements of panelists with specific knowledge of the disadvantages faced by individuals poorly equipped to cope with the reading demands they commonly confront. Panelists were asked to evaluate a large number of draft tasks, to recommend revision or deletion where necessary, and to suggest important classes of tasks not represented in the draft collection. For a variety of reasons

having largely to do with the structure of the meetings and the time available, the main outcomes of these meetings were identification of general principles rather than a set of specific inclusion-exclusion decisions. Despite some lack of success here, however, the panels did make a number of recommendations which were the basis for further development and revision.

The product of these efforts at this stage is a collection of about 200 tasks that are judged to reflect faithfully reading demands encountered in a variety of practical settings. These tasks will be tried out on adult samples in the near future and the results of these and subsequent administrations of the tasks should provide a great deal of information about the ability of adults to cope with these demands. The utility of the pool of tasks as a resource for the construction of literacy assessment instruments should be evidenced by the results of these trials and by subsequent uses of smaller parts of the collection.